The Curse of Dimensionality for Numerical Integration of Smooth Functions

Aicke Hinrichs, Erich Novak*
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, 07743 Jena, Germany
email: a.hinrichs@uni-jena.de, erich.novak@uni-jena.de

Mario Ullrich†
Dipartimento di Matematica, Università Roma Tre
Largo San Leonardo Murialdo 1, 00146 Roma, Italy
email: ullrich.mario@gmail.com

Henryk Woźniakowski‡
Department of Computer Science, Columbia University,
New York, NY 10027, USA, and
Institute of Applied Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warszawa, Poland
email: henryk@cs.columbia.edu

November 6, 2012

*This author was partially supported by the DFG-Priority Program 1324.
†This author was supported by DFG GRK 1523 and ERC Advanced Grant PTRELSS.
‡This author was partially supported by the National Science Foundation.

Abstract

We prove the curse of dimensionality for multivariate integration of C^k functions. The proofs are based on volume estimates for $k = 1$ together with smoothing by convolution. This allows us to obtain smooth fooling functions for $k > 1$.

MSC: 65D30,65Y20,41A63,41A55

Keywords: curse of dimensionality, numerical integration, high dimensional numerical problems

1 Introduction

We study multivariate integration for different classes F_d of smooth functions $f \colon \mathbb{R}^d \to \mathbb{R}$. Our emphasis is on large values of $d \in \mathbb{N}$. We want to approximate

$$S_d(f) = \int_{D_d} f(x)\,\mathrm{d}x \quad \text{for} \quad f \in F_d \tag{1}$$

up to some error $\varepsilon > 0$, where $D_d \subset \mathbb{R}^d$ has (Lebesgue) measure 1. The results in this paper hold for arbitrary sets D_d, the standard example of course is $D_d = [0,1]^d$.

We consider (deterministic) algorithms that use only function values. We consider classes F_d of functions bounded in absolute value by 1 and containing all constant functions $f(x) \equiv c$ with $|c| \le 1$. This implies that the initial error is one, i.e.,

$$\inf_{c \in \mathbb{R}} \max_{f \in F_d} |S_d(f) - c| = \max_{f \in F_d} |S_d(f)| = 1,$$

so that multivariate integration is well scaled and that is why we consider $\varepsilon < 1$.

Let $n(\varepsilon, F_d)$ denote the minimal number of function values needed for this task in the worst case setting[1]. By the *curse of dimensionality* we mean that $n(\varepsilon, F_d)$ is exponentially large in d. That is, there are positive numbers c, ε_0 and γ such that

$$n(\varepsilon, F_d) \ge c\,(1+\gamma)^d \quad \text{for all} \quad \varepsilon \le \varepsilon_0 \quad \text{and infinitely many} \quad d \in \mathbb{N}. \tag{2}$$

For many natural classes F_d the bound in (2) will hold for all $d \in \mathbb{N}$. This applies in particular to the classes considered in this paper.

[1] We add that $n(\varepsilon, F_d)$ is the information complexity of multivariate integration over F_d and is proportional to the (total) complexity as long as F_d is convex and symmetric. The last two assumptions are needed to guarantee that a linear algorithm is optimal and its implementation cost is linear in $n(\varepsilon, F_d)$.

There are many classes F_d for which the curse of dimensionality has been proved, see [5, 7] for such examples. However, it has been *not* known if the curse of dimensionality occurs for probably the most natural class which is the unit ball of r times continuously differentiable functions,

$$\mathcal{C}_d^r = \{f \in C^r(\mathbb{R}^d) \mid \ \|D^\beta f\| \le 1 \quad \text{for all} \quad |\beta| \le r\},$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_d)$, with non-negative integers β_j, $|\beta| = \sum_{j=1}^d \beta_j$, and D^β denotes the operator of β_j times differentiation with respect to the jth variable for $j = 1, 2, \dots d$. By $\|\cdot\|$ we mean the sup norm, $\|D^\beta f\| = \sup_{x\in\mathbb{R}^d} |(D^\beta f)(x)|$.

For $r = 0$, we obviously have $n(\varepsilon, \mathcal{C}_d^0) = \infty$ for all $\varepsilon < 1$ and all $d \in \mathbb{N}$. Therefore from now on we always assume that $r \ge 1$. For $r = 1$, the curse of dimensionality for $\mathcal{C}_d^1$ follows from the results of Sukharev [8]. Whether the curse holds for $r \ge 2$ has been an open problem for many years.

The class $\mathcal{C}_d^r$ for $D_d = [0,1]^d$ (and functions and norms restricted to D_d) was already studied in 1959 by Bakhvalov [2], see also [4]. He proved that there are two positive numbers $a_{d,r}$ and $A_{d,r}$ such that

$$a_{d,r}\, \varepsilon^{-d/r} \le n(\varepsilon, \mathcal{C}_d^r) \le A_{d,r}\, \varepsilon^{-d/r} \quad \text{for all } d \in \mathbb{N} \text{ and } \varepsilon \in (0,1). \tag{3}$$

This means that for a fixed d and for ε tending to zero, we know that $n(\varepsilon, \mathcal{C}_d^r)$ is of order $\varepsilon^{-d/r}$ and the exponent of ε^{-1} grows linearly[2] in d. Unfortunately, since the known dependence on d in $a_{d,r}$ is exponentially small and the known dependence on d in $A_{d,r}$ is exponentially large in d, Bakhvalov's result does not allow us to conclude whether the curse of dimensionality holds for the class $\mathcal{C}_d^r$. In fact, if we reverse the roles of d and ε, and consider a fixed ε and d tending to infinity, the bound (3) on $n(\varepsilon, \mathcal{C}_d^r)$ is useless. We prove the following result.

Main Theorem. The curse of dimensionality holds for the classes $\mathcal{C}_d^r$ with the *super-exponential* lower bound

$$n(\varepsilon, \mathcal{C}_d^r) \ \ge\ c_r\, (1-\varepsilon)\, d^{\,d/(2r+3)} \quad \text{for all } d \in \mathbb{N} \text{ and } \varepsilon \in (0,1),$$

where $c_r \in (0,1]$ depends only on r.

We also prove that the curse of dimensionality holds for even smaller classes of functions F_d for which the norms of arbitrary directional derivatives are bounded proportionally to $1/\sqrt{d}$.

[2]In the language of tractability, this result means that we do *not have* polynomial tractability but does not allow us to conclude the lack of weak tractability.

We now discuss how we obtain lower bounds on $n(\varepsilon, F_d)$ for numerical integration defined on convex and symmetric classes F_d. The standard proof technique is to find a fooling function $f \in F_d$ that vanishes at the points $\mathcal{P} = \{x_1, x_2, \dots, x_n\}$ at which we sample functions from F_d, and the integral of f is as large as possible. All algorithms that use function values at x_j's must give the same approximation of the integral of f and of the integral of $-f$. That is why the integral of f is a lower bound on the worst case error of all algorithms using function values at x_j's. If for all choices of $x_1, x_2, \dots, x_n$ the integral of f is larger than ε then we know that $n(\varepsilon, F_d) \geq n$.

We start with the fooling function

$$f_0(x) = \min\left\{1, \frac{1}{\delta\sqrt{d}}\,\mathrm{dist}(x, \mathcal{P}_\delta)\right\} \quad \text{for all} \quad x \in \mathbb{R}^d,$$

where

$$\mathcal{P}_\delta = \bigcup_{i=1}^{n} B_\delta^d(x_i)$$

and $B_\delta^d(x_i)$ is the ball with center x_i and radius $\delta\sqrt{d}$. The function f_0 is Lipschitz. By a suitable smoothing via convolution we construct a fooling function $f_r \in \mathcal{C}_d^r$ and $f_r|_{\mathcal{P}} = 0$.

2 Preliminaries

In this section we precisely define our problem. Let F_d be a class of Lebesgue integrable functions $f : \mathbb{R}^d \to \mathbb{R}$. For $f \in F_d$, we approximate the integral $S_d(f)$, see (1), by algorithms

$$A_{n,d}(f) = \phi_{n,d}\big(f(x_1), f(x_2), \dots, f(x_n)\big),$$

where $x_j \in \mathbb{R}^d$ can be chosen adaptively and $\phi_{n,d} : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary mapping. Adaption means that the selection of x_j may depend on the already computed values $f(x_1), f(x_2), \dots, f(x_{j-1})$. The (worst case) error of the algorithm $A_{n,d}$ is defined as

$$e(A_{n,d}) = \sup_{f \in F_d} |S_d(f) - A_{n,d}(f)|.$$

The minimal number of function values to guarantee that the error is at most ε is defined as

$$n(\varepsilon, F_d) = \min\{\, n \in \mathbb{N} \;\big|\; \exists\, A_{n,d} \text{ such that } e(A_{n,d}) \leq \varepsilon\}.$$

Hence we minimize n over all choices of adaptive sample points x_j and mappings $\phi_{n,d}$. It is well known that as long as the class F_d is convex and symmetric we may restrict the

minimization of n by considering only nonadaptive choices of x_j and linear mappings $\phi_{n,d}$. Furthermore,

$$n(\varepsilon, F_d) = \min\Big\{\, n \in \mathbb{N} \mid \inf_{\mathcal{P}\subset\mathbb{R}^d, \#\mathcal{P}=n} \ \sup_{f\in F_d, f|_{\mathcal{P}}=0} |S_d(f)| \le \varepsilon \Big\}, \tag{4}$$

see [4, Prop. 1.2.6] or [9, Theorem 5.5.1]. In this paper we always consider convex and symmetric F_d so that we can use the last formula for $n(\varepsilon, F_d)$. For more details see, e.g., Chapter 4 in [5].

Observe that we allow $x_j \in \mathbb{R}^d$ instead of only $x_j \in D_d$. In this paper we are interested in *lower* bounds and this assumption makes our results even stronger.

As already mentioned, our lower bounds are based on a volume estimate of a neighborhood of certain sets in $\mathbb{R}^d$, see also [3]. In the following we denote by A_δ, for $A \subset \mathbb{R}^d$, the $(\delta\sqrt{d})$-neighborhood of A, which is defined by

$$A_\delta = \big\{ x \in \mathbb{R}^d \mid \operatorname{dist}(x, A) \le \delta\sqrt{d} \big\}, \tag{5}$$

where $\operatorname{dist}(x, A) = \inf_{a\in A} \|x - a\|_2$ denotes the Euclidean distance of x from A.

Since we need the $\sqrt{d}$-scaling of the distance, we will omit it in the notation as we already did for A_δ. Furthermore, we denote by $B_\delta^d(x)$ the d-dimensional ball with center $x \in \mathbb{R}^d$ and radius $\delta\sqrt{d}$, i.e.,

$$B_\delta^d(x) \;=\; \big\{ y \in \mathbb{R}^d \mid \ \|x - y\|_2 \le \delta\sqrt{d} \big\}.$$

We will need some standard volume estimates for Euclidean balls. Recall that the volume of a Euclidean ball of radius 1 is given by

$$V_d = \frac{\pi^{d/2}}{\Gamma(1 + d/2)}.$$

From Stirling's formula for the Γ function we have

$$\Gamma(x + 1) = \sqrt{2\pi x}\, x^x\, \mathrm{e}^{-x+\frac{\theta_x}{12x}} \quad \text{for all} \quad x > 0,$$

where $\theta_x \in (0, 1)$, see [1, p. 257]. This leads to the estimate

$$\Gamma(x + 1) > \sqrt{2\pi x} \left(\frac{x}{\mathrm{e}}\right)^x \quad \text{for all} \quad x > 0.$$

Combining this estimate with the volume formula for the ball, we obtain for all $d \in \mathbb{N}$,

$$\lambda_d\big(B_\delta^d(x)\big) = \big(\delta\sqrt{d}\big)^d V_d < \big(\delta\sqrt{d}\big)^d \frac{\left(\frac{2\pi\mathrm{e}}{d}\right)^{d/2}}{\sqrt{\pi d}} = \frac{\big(\delta\sqrt{2\pi\mathrm{e}}\big)^d}{\sqrt{\pi d}} < \Big(\delta\sqrt{2\pi\mathrm{e}}\Big)^d. \tag{6}$$

The volume formula for the Euclidean unit ball also shows the recurrence relation

$$\frac{V_{d-1}}{V_d} = \frac{d}{d-1}\,\frac{V_{d-3}}{V_{d-2}} \quad \text{for all} \quad d \geq 4.$$

This easily implies

$$\frac{2}{\sqrt{d}}\,\frac{V_{d-1}}{V_d} < \frac{2}{\sqrt{d-2}}\,\frac{V_{d-3}}{V_{d-2}} \quad \text{for all} \quad d \geq 4.$$

The last inequality can be used in an inductive argument leading to

$$\frac{2}{\sqrt{d}}\,\frac{V_{d-1}}{V_d} \leq 1 \quad \text{for all} \quad d \geq 2. \tag{7}$$

This will be needed later.

3 Convolution

In this section we fix $k \in \mathbb{N}$ and study the convolution

$$f_k := f * g_1 * \ldots * g_k$$

of a function f defined on $\mathbb{R}^d$ with (normalized) indicator functions g_j. We are interested in properties of f_k in terms of the properties of the initial function f. Recall that the convolution of two functions f and g on $\mathbb{R}^d$ is defined by

$$(f * g)(x) = \int_{\mathbb{R}^d} f(x-t)\, g(t)\, \mathrm{d}t \quad \text{for all} \quad x \in \mathbb{R}^d.$$

Fix a number $\delta > 0$ and a sequence $(\alpha_j)_{j=1}^k$ with $\alpha_j > 0$ such that

$$\sum_{j=1}^{k} \alpha_j \leq 1.$$

For example, we may take $\alpha_j = 1/k$ for $j = 1, 2 \ldots, k$. For $j = 1, \ldots, k$, we define the ball

$$B_j = \left\{ x \in \mathbb{R}^d \,\middle|\, \|x\|_2 \leq \alpha_j\, \delta \sqrt{d} \right\}$$

and the function $g_j \colon \mathbb{R}^d \to \mathbb{R}$ by

$$g_j(x) = \frac{\mathbb{1}_{B_j}(x)}{\lambda_d(B_j)} = \frac{1}{\lambda_d(B_j)} \begin{cases} 1 & \text{if } x \in B_j, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Thus, the convolution of a function f with g_j can be written as

$$(f * g_j)(x) \;=\; \frac{1}{\lambda_d(B_j)} \int_{B_j} f(x+t)\,\mathrm{d}t \quad \text{for all} \quad x \in \mathbb{R}^d.$$

We will frequently use the following probabilistic interpretation. Let Y_j be a random variable that is uniformly distributed on B_j. Then the convolution of f with g_j can be written as the expected value

$$(f * g_j)(x) \;=\; \mathbb{E}\big[f(x+Y_j)\big].$$

The next theorem is the basis for the induction steps of the proofs of our main results. For $f\colon \mathbb{R}^d \to \mathbb{R}$, we use the Lipschitz constant

$$\mathrm{Lip}(f) = \sup_{x \neq y} \frac{|f(x)-f(y)|}{\|x-y\|_2}.$$

Define

$$C^r = \big\{f\colon \mathbb{R}^d \to \mathbb{R} \mid \quad D^{\theta_\ell} \dots D^{\theta_1} f \text{ is continuous for all } \ell \le r \text{ and all } \theta_1, \dots, \theta_r \in \mathbb{S}^{d-1}\big\},$$

where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$ and $D^{\theta_1} f(x) = \lim_{h \to 0} \frac{1}{h}\big(f(x+h\theta_1) - f(x)\big)$ is the derivative in the direction of θ_1.

Theorem 1. For $k \in \mathbb{N}$ and $f \in C^r$, define

$$f_k = f * g_1 * \dots * g_k \quad \text{with} \;\; g_k \text{ from (8).}$$

For $d \ge 2$, let $\Omega \subset \mathbb{R}^d$ and let Ω_δ be its neighborhood defined as in (5). Then

(i) if $f(x) = 0$ for all $x \in \Omega_\delta$ then $f_k(x) = 0$ for all $x \in \Omega$,

(ii) $\mathrm{Lip}(f_k) \le \mathrm{Lip}(f)$,

(iii) if $\int_{D_d} f(x+t)\,\mathrm{d}x \;\ge\; \varepsilon$ for all $t \in \mathbb{R}^d$ with $\|t\|_2 \le \delta\sqrt{d}$ then $\int_{D_d} f_k(x)\mathrm{d}x \ge \varepsilon$,

(iv) for all $\ell \le r$ and all $\theta_1, \theta_2, \dots, \theta_r \in \mathbb{S}^{d-1}$,

$$\mathrm{Lip}\Big(D^{\theta_\ell}\, D^{\theta_{\ell-1}} \dots D^{\theta_1} f_k\Big) \le \mathrm{Lip}\Big(D^{\theta_\ell}\, D^{\theta_{\ell-1}} \dots D^{\theta_1} f\Big),$$

(v) $f_k \in C^{r+k}$, and for all $\ell \le r$, all $j = 1, \dots, k$ and all $\theta_1, \theta_2, \dots, \theta_{\ell+j} \in \mathbb{S}^{d-1}$,

$$\mathrm{Lip}\Big(D^{\theta_{\ell+j}}\, D^{\theta_{\ell+j-1}} \dots D^{\theta_1} f_k\Big) \le \left(\prod_{i=1}^{j} \frac{1}{\delta \alpha_{k+1-i}}\right) \mathrm{Lip}\Big(D^{\theta_\ell}\, D^{\theta_{\ell-1}} \dots D^{\theta_1} f\Big).$$

The parts (i)–(iv) of this theorem show that some properties of the initial function f are preserved by convolutions. Part (v) states that we gain one "degree of smoothness" with every convolution, loosing only a multiplicative constant for its Lipschitz constant.

Proof. First note that we can write f_k as

$$f_k(x) \,=\, \mathbb{E}\big[f(x+Y)\big], \quad \text{for all} \quad x \in \mathbb{R}^d,$$

where Y is a random variable with probability density function $g_1 * \ldots * g_k$. By construction of g_k's which are the indicator functions of the balls whose sum of the radii is at most $\delta\sqrt{d}$, we have

$$\{t \in \mathbb{R}^d \mid \; g_1 * \ldots * g_k(t) > 0\} \subset \{t \in \mathbb{R}^d \mid \; \|t\|_2 \le \delta\sqrt{d}\},$$

which implies that $x + Y \in \Omega_\delta$ almost surely for every $x \in \Omega$. Thus, $f(x) = 0$ for all $x \in \Omega_\delta$ implies that $f_k(x) = 0$ for all $x \in \Omega$, which is property (i).

Property (ii) is proven by

$$\begin{aligned} |f_k(x) - f_k(y)| \,&=\, \big|\mathbb{E}\big[f_k(x+Y) - f_k(y+Y)\big]\big| \,\le\, \mathbb{E}\big[|f(x+Y) - f(y+Y)|\big] \\ &\le\, \mathrm{Lip}(f)\,\mathbb{E}\big[\|(x+Y)-(y+Y)\|_2\big] \,=\, \mathrm{Lip}(f)\,\|x-y\|_2\,. \end{aligned}$$

To prove (iii), we use Fubini's theorem and we obtain

$$\int_{D_d} f_k(x)\,\mathrm{d}x \,=\, \int_{D_d} \mathbb{E}\big[f(x+Y)\big]\,\mathrm{d}x \,=\, \mathbb{E}\Big[\int_{D_d} f(x+Y)\,\mathrm{d}x\Big] \,\ge\, \varepsilon$$

by assumption.

For the proof of properties (iv) and (v), let $\theta = (\theta_1, \ldots, \theta_\ell) \in (\mathbb{S}^{d-1})^\ell$. We write D^θ for $D^{\theta_\ell} \ldots D^{\theta_1}$. Clearly, $f \in C^r$ and $\ell \le r$ implies that $D^\theta f \in C^{r-\ell} \subseteq C$. Since f_k is at least as smooth as f, both $D^\theta f$ and $D^\theta f_k$ are well defined.

We need the well-known fact that $D^\theta(f * g) = (D^\theta f) * g$ if $f \in C^\ell$ and g has compact support. For $g = g_1 * \ldots * g_k$, we have

$$\begin{aligned} \big|D^\theta f_k(x) - D^\theta f_k(y)\big| \,&=\, \big|\big((D^\theta f) * g\big)(x) - \big((D^\theta f) * g\big)(y)\big| \\ &=\, \left|\int_{\mathbb{R}^d} \big[(D^\theta f(x+t) - D^\theta f(y+t)\big]\; g(t)\mathrm{d}t\right| \\ &\le\, \mathrm{Lip}(D^\theta f)\;\|x-y\|_2 \int_{\mathbb{R}^d} g(t)\mathrm{d}t \\ &=\, \mathrm{Lip}(D^\theta f)\;\|x-y\|_2 \end{aligned}$$

for all $x, y \in \mathbb{R}^d$. The last equality follows since the g_k is normalized. This proves (iii).

For (v), we need to prove that $f_k \in C^{r+k}$ with $f_0 = f \in C^r$ by assumption, and then it is enough to show that for all $m \le r+k$ and all $\theta = (\theta_m, \dots, \theta_1) \in (\mathbb{S}^{d-1})^m$,

$$\operatorname{Lip}\Big(D^\theta f_k\Big) \le \frac{1}{\delta \alpha_k} \operatorname{Lip}\Big(D^{\bar\theta} f_{k-1}\Big),$$

where $\bar\theta = (\theta_{m-1}, \dots, \theta_1) \in (\mathbb{S}^{d-1})^{m-1}$.

Assume inductively that $f_{k-1} \in C^{m-1}$, which holds for $k = 1$. This implies $D^{\bar\theta}(f_{k-1} * g_k) = (D^{\bar\theta} f_{k-1}) * g_k$, and

$$\begin{aligned}
D^\theta f_k(x) \;&=\; D^{\theta_m}\big((D^{\bar\theta} f_{k-1}) * g_k\big)(x) \\
&=\; D^{\theta_m}\Big(\frac{1}{\lambda_d(B_k)} \int_{\mathbb{R}^d} D^{\bar\theta} f_{k-1}(x+t)\, \mathbb{1}_{B_k}(t)\,\mathrm{d}t\Big) \\
&=\; \frac{1}{\lambda_d(B_k)}\, D^{\theta_m}\Big(\int_{\theta_m^\perp} \int_{\mathbb{R}} D^{\bar\theta} f_{k-1}(x+s+h\theta_m)\, \mathbb{1}_{B_k}(s+h\theta_m)\,\mathrm{d}h\,\mathrm{d}s\Big) \\
&=\; \frac{1}{\lambda_d(B_k)} \int_{\theta_m^\perp} D^{\theta_m}\Big(\int_{\mathbb{R}} D^{\bar\theta} f_{k-1}(x+s+h\theta_m)\, \mathbb{1}_{B_k}(s+h\theta_m)\,\mathrm{d}h\Big)\,\mathrm{d}s,
\end{aligned}$$

where $\theta_m^\perp$ is the hyperplane orthogonal to θ_m. For any function f on $\mathbb{R}$ of the form

$$f(x) = \int_{x-a}^{x+a} g(y)\,\mathrm{d}y$$

with some continuous function g we have

$$f'(x) = g(x+a) - g(x-a).$$

Therefore, we obtain

$$\begin{aligned}
D^\theta f_k(x) \;=\; \frac{1}{\lambda_d(B_k)} \int_{B_k \cap \theta_m^\perp} & D^{\bar\theta} f_{k-1}\Big(x+s+h_{\max}(s)\,\theta_m\Big) \\
& - D^{\bar\theta} f_{k-1}\Big(x+s-h_{\max}(s)\,\theta_m\Big)\,\mathrm{d}s
\end{aligned}$$

with

$$h_{\max}(s) \;=\; \max\{h \ge 0 \mid \quad s + h\theta_m \in B_k\}.$$

For each $s \in B_k \cap \theta_m^\perp$, define the points $s_1 = s + h_{\max}(s)\,\theta_m \in B_k$ and

$s_2 = s - h_{\max}(s)\,\theta_m \in B_k$. Then

$$\begin{aligned}
\left|D^\theta f_k(x) - D^\theta f_k(y)\right| &\le \frac{1}{\lambda_d(B_k)} \int_{B_k \cap \theta_m^\perp} \Big[\Big|D^{\bar\theta} f_{k-1}\big(x+s_1\big) - D^{\bar\theta} f_{k-1}\big(x+s_2\big) \\
&\qquad\qquad - D^{\bar\theta} f_{k-1}\big(y+s_1\big) + D^{\bar\theta} f_{k-1}\big(y+s_2\big)\Big|\Big]\,\mathrm{d}s \\
&\le \frac{1}{\lambda_d(B_k)} \int_{B_k \cap \theta_m^\perp} \Big[\Big|D^{\bar\theta} f_{k-1}\big(x+s_1\big) - D^{\bar\theta} f_{k-1}\big(y+s_1\big)\Big| \\
&\qquad\qquad + \Big|D^{\bar\theta} f_{k-1}\big(x+s_2\big) - D^{\bar\theta} f_{k-1}\big(y+s_2\big)\Big|\Big]\,\mathrm{d}s \\
&\le \frac{2\,\lambda_{d-1}(B_k \cap \theta_m^\perp)}{\lambda_d(B_k)}\, \mathrm{Lip}(D^{\bar\theta} f_{k-1})\, \|x-y\|_2\,.
\end{aligned}$$

In particular, this shows the implication

$$f_{k-1} \in C^{m-1} \implies f_k \in C^m$$

for all $k \in \mathbb{N}$. Taking $m = r + k$ we have $f_k \in C^{r+k}$, as claimed.

For $m \le r+k$, it remains to bound $2\lambda_{d-1}(B_k \cap \theta_m^\perp)/\lambda_d(B_k)$. Recall that B_k is a ball with radius $\delta\alpha_k\sqrt{d}$ and that V_d is the volume of the Euclidean unit ball in $\mathbb{R}^d$. We obtain from (7) that

$$\frac{2\,\lambda_{d-1}(B_k \cap \theta_m^\perp)}{\lambda_d(B_k)} = \frac{2(\delta\alpha_k\sqrt{d})^{d-1}}{(\delta\alpha_k\sqrt{d})^d}\,\frac{V_{d-1}}{V_d} = \frac{2}{\delta\alpha_k\sqrt{d}}\,\frac{V_{d-1}}{V_d} \le \frac{1}{\delta\alpha_k}.$$

□

4 Main Results

Let $\mathcal{P} = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ be a collection of n points. As pointed out in the introduction, we want to construct functions that vanish at $\mathcal{P}$ and have a large integral. For this, we choose

$$f_0(x) = \min\left\{1, \frac{1}{\delta\sqrt{d}}\,\mathrm{dist}(x, \mathcal{P}_\delta)\right\} \quad \text{for all} \quad x \in \mathbb{R}^d,$$

where

$$\mathcal{P}_\delta = \bigcup_{i=1}^n B_\delta^d(x_i)$$

and $B_\delta^d(x_i)$ is the ball with center x_i and radius $\delta\sqrt{d}$.

The function $\mathrm{dist}(\cdot, \mathcal{P}_\delta)$ is Lipschitz with constant 1. Hence for $\delta \le 1$

$$\mathrm{Lip}(f_0) = \frac{1}{\delta\sqrt{d}}. \tag{9}$$

Additionally, $f_0(x) = 0$ for all $x \in \mathcal{P}_\delta$ by definition.

Using these facts we can apply Theorem 1 to prove the curse of dimensionality for the following class of functions that are defined on $\mathbb{R}^d$. For a fixed $r \in \mathbb{N}$, we now take $\alpha_1 = \cdots = \alpha_r = \frac{1}{r}$ and define

$$F_{d,r,\delta} = \{f\colon \mathbb{R}^d \to \mathbb{R} \mid \ f \in C^r \text{ satisfies (10)–(12)}\},$$

where

$$\begin{aligned} \|f\| &\le 1, && (10)\\ \mathrm{Lip}(f) &\le \frac{1}{\delta\sqrt{d}}, && (11)\\ \forall k \le r: \max_{\theta_1,\dots,\theta_k \in \mathbb{S}^{d-1}} \mathrm{Lip}(D^{\theta_1} \dots D^{\theta_k} f) &\le \frac{1}{\delta\sqrt{d}} \left(\frac{r}{\delta}\right)^k. && (12) \end{aligned}$$

Theorem 2. For any $r \in \mathbb{N}$ and $\delta \in (0,1]$,

$$n(\varepsilon, F_{d,r,\delta}) \ \ge \ (1-\varepsilon) \begin{cases} 1 & \text{for} \quad d = 1, \\ \left(\delta\sqrt{18e\pi}\right)^{-d} & \text{for} \quad d \ge 2, \end{cases} \qquad \text{for all } \varepsilon \in (0,1).$$

Hence the curse of dimensionality holds for the class $F_{d,r,\delta}$ for $\delta < 1/\sqrt{18e\pi}$.

Note that this result shows that the growth rate of $n(\varepsilon, F_{d,r,\delta})$ in d can be arbitrarily large if we choose δ small enough.

Proof. Since the initial error for the classes $F_{d,r,\delta}$ is 1 we obtain $n(\varepsilon, F_{d,r,\delta}) \ge 1$ for all $\varepsilon \in (0,1)$. This proves the statement for $d = 1$.

For $d \ge 2$, we use Theorem 1 with $k = r$, $\Omega = \mathcal{P}$ and $f_r(x) = f_0 * g_1 * \ldots * g_r(x)$. Here g_k's are as in Theorem 1. Recall that we have chosen $\alpha_1 = \ldots = \alpha_r = 1/r$ and $\alpha_j = 0$ for $j > r$. The properties of the initial function f_0 and Theorem 1 immediately imply that f_r satisfies (10)–(12). It remains to bound its integral. Note that $f_0(x) = 1$ for all $x \notin \mathcal{P}_{2\delta}$. Clearly, $f_r(x) \ge 0$ for all $x \in \mathbb{R}^d$. Since $f_r(x)$ depends only on the values $f_0(x+t)$ for $t \in \mathbb{R}^d$

with $\|t\|_2 \le \delta\sqrt{d}$, it follows that $f_r(x) = 1$ for $x \notin \mathcal{P}_{3\delta}$. We thus obtain

$$\begin{aligned}\int_{D_d} f_r(x)\,\mathrm{d}x \;&\ge\; \int_{D_d\setminus\mathcal{P}_{3\delta}} f_r(x)\,\mathrm{d}x \;=\; 1-\lambda_d(\mathcal{P}_{3\delta}\cap D_d)\\ &\ge\; 1-\lambda_d(\mathcal{P}_{3\delta}) \;\ge\; 1-n\lambda_d(B^d_{3\delta})\\ &>\; 1-\frac{n\Big(3\delta\sqrt{2e\pi}\Big)^d}{\sqrt{\pi d}}\\ &>\; 1-n\Big(3\delta\sqrt{2e\pi}\Big)^d,\end{aligned}$$

where the last inequality follows from the bound that is given in (6). Hence $\int_{D_d} f_r(x)\,\mathrm{d}x \le \varepsilon$ implies that

$$n \ge (1-\varepsilon)\,(\delta\sqrt{18\varepsilon\pi})^{-d}.$$

Since this holds for arbitrary $\mathcal{P}$, the result follows. $\square$

By Theorem 2, we know how the parameter δ comes into play. For $p > 0$, let

$$\delta \;=\; \frac{1}{\sqrt{18e\pi}}\, d^{-p/(r+1)}.$$

For this δ, we obtain a somehow stronger form of the curse of dimensionality for the class

$$\widetilde{F}_{d,r,p} \;=\; \{f\colon \mathbb{R}^d \to \mathbb{R} \;\big|\; f \in C^r \text{ satisfies (13)–(15)}\},$$

where

$$\begin{aligned}\|f\| \;&\le\; 1, && (13)\\ \mathrm{Lip}(f) \;&\le\; d^{-\frac{1}{2}+\frac{p}{r+1}}\,\sqrt{18e\pi}, && (14)\\ \forall k \le r:\; \max_{\theta_1,\dots,\theta_k\in\mathbb{S}^{d-1}} \mathrm{Lip}(D^{\theta_1}\dots D^{\theta_k} f) \;&\le\; d^{-\frac{1}{2}+\frac{p(k+1)}{r+1}}\, r^k \left(\sqrt{18e\pi}\right)^{k+1}. && (15)\end{aligned}$$

Theorem 3. For any $r \in \mathbb{N}$ and $p > 0$,

$$n(\varepsilon, \widetilde{F}_{d,r,p}) \;\ge\; (1-\varepsilon)\, d^{\,p\,d/(r+1)} \quad \text{for all } d \in \mathbb{N} \text{ and } \varepsilon \in (0,1).$$

Hence the curse of dimensionality holds for the class $\widetilde{F}_{d,r,p}$.

Note that the classes $\widetilde{F}_{d,r,p}$ are contained in the classes

$$\mathcal{C}_d^r = \{f \in C^r \mid \|D^\beta f\| \le 1 \quad \text{for all} \quad |\beta| \le r\},$$

if $p < 1/2$ and d is large enough. This holds if

$$d \ \ge \ \left(r^r \, (18 \mathrm{e} \pi)^{(r+1)/2}\right)^{1/(1/2-p)}. \tag{16}$$

From this we easily obtain the main result already stated in the introduction.

Main Theorem. For any $r \in \mathbb{N}$, there exists a constant $c_r \in (0,1]$ such that

$$n(\varepsilon, \mathcal{C}_d^r) \ge c_r \, (1-\varepsilon) \, d^{\, d/(2r+3)} \qquad \text{for all } \ d \in \mathbb{N} \ \text{ and } \ \varepsilon \in (0,1).$$

Hence the curse of dimensionality holds for the class $\mathcal{C}_d^r$.

Proof. The case $d = 1$ is trivial since the initial error for the classes $\mathcal{C}_d^r$ is again 1.

For $d \ge 2$, we know from Theorem 3 and the discussion thereafter that $n(\varepsilon, \mathcal{C}_d^r) \ge (1-\varepsilon) \, d^{pd/(r+1)}$ for all $p < 1/2$ if $d \ge d_0$, where $d_0 = d_0(r,p)$ is the right hand side of (16). This implies that for

$$\widetilde{c}_{r,p} = d_0^{-pd_0/(r+1)},$$

which depends only on r and p, we have

$$n(\varepsilon, \mathcal{C}_d^r) \ge \widetilde{c}_{r,p} \, (1-\varepsilon) \, d^{pd/(r+1)} \quad \text{for all} \quad d \ge 2.$$

The choice $p^* = (r+1)/(2r+3)$ yields the result with $c_r = \widetilde{c}_{r,p^*}$. □

Note that c_r in the last theorem is super exponentially small in r.

Remark 1. The reader might find it more natural to define classes of functions $F_{d,r}(D_d)$ that are defined only on $D_d \subset \mathbb{R}^d$. Not all such functions can be extended to smooth functions on $\mathbb{R}^d$, and even if they can be extended then the norm of the extended function could be much larger. Our lower bound results for functions defined on $\mathbb{R}^d$ can be also applied for functions defined on $D_d \subset \mathbb{R}^d$ and this makes them even stronger.

Remark 2. Note that the possibility of super-exponential lower bounds on the complexity depends on the definition of the Lipschitz constant. For the class

$$F_d = \left\{ f \colon [0,1]^d \to \mathbb{R} \mid \sup_{x,y \in [0,1]^d} \frac{|f(x) - f(y)|}{\|x-y\|_\infty} \le 1 \right\},$$

Sukharev [8] proved that for $n = m^d$ the midpoint rule is optimal with error $e_n = \frac{d}{2d+2} n^{-1/d}$. Hence, roughly, $n(\varepsilon, F_d) \approx 2^{-d} \varepsilon^{-d}$ and the complexity is "only" exponential in d for $\varepsilon < 1/2$.

Remark 3. We mention two results for the very small class

$$F_d = C_d^\infty = \{f \in C^\infty([0,1]^d) \mid \quad \|D^\beta f\| \le 1 \quad \text{for all} \quad \beta \in \mathbb{N}_0^d\}.$$

O. Wojtaszczyk [10] proved that $\lim_{d\to\infty} n(\varepsilon, F_d) = \infty$ for every $\varepsilon < 1$, hence the problem is not strongly polynomially tractable. It is still open whether the curse of dimensionality holds for this class F_d. The same class F_d was studied for the approximation problem in [6]. For this problem the curse of dimensionality is present even if we allow algorithms that use arbitrary linear functionals.

References

[1] M. Abramowitz, I. A. Stegun, *Handbook of mathematical functions*, tenth printing, NBS, Applied Mathematics Series 55, 1972.

[2] N. S. Bakhvalov, On approximate computation of integrals, *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem*, **4**, 3–18, 1959, in Russian.

[3] A. Hinrichs, E. Novak, H. Woźniakowski, The curse of dimensionality for the class of monotone functions and for the class of convex functions, *J. Approx. Th.* **163**, 955–965, 2011.

[4] E. Novak, *Deterministic and Stochastic Error Bounds in Numerical Analysis*, LNiM **1349**, Springer-Verlag, Berlin, 1988.

[5] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, Volume I: Linear Information, European Math. Soc. Publ. House, Zürich, 2008.

[6] E. Novak and H. Woźniakowski, Approximation of infinitely differentiable multivariate functions is intractable, *J. Complexity* **25**, 398–404, 2009.

[7] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems*, Volume II: Standard Information for Functionals, European Math. Soc. Publ. House, Zürich, 2010.

[8] A. G. Sukharev, Optimal numerical integration formulas for some classes of functions of several variables, *Soviet Math. Dokl.* **20**, 472–475, 1979.

[9] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski, *Information-Based Complexity*, Academic Press, 1988.

[10] O. Wojtaszczyk, Multivariate integration in $C^\infty([0,1]^d)$ is not strongly tractable, *J. Complexity* **19**, 638–643, 2003.